(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/02088

(22) International Filing Date: 18 January 2001 (18.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/177,719   21 January 2000 (21.01.2000)   US
60/178,988   28 January 2000 (28.01.2000)   US
60/184,959   25 February 2000 (25.02.2000)   US
60/190,112   17 March 2000 (17.03.2000)   US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94522 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). BAUGHN, Mariah, R. [US/US]; 14211 Santiago Road, San Leandro, CA 94577 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #17, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). WANG, Eureka

[US/US]; 4621 Strawberry Park Drive, San Jose, CA 95129 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). POLICKY, Jennifer, L. [US/US]; 1861 National Avenue, San Jose, CA 95124 (US). MATHUR, Preete [US/US]; 43733 Greenhills Way, Fremont, CA 94539 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
    without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/53469 A2

PHOSPHATASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of phosphatases and to the use

5   of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological

disorders, developmental disorders and cell proliferative disorders, including cancer, and in the

assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid

sequences of phosphatases.

10   ## BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the

intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in

a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate

groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues

15   serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups.

Extracellular signals including hormones, neurotransmitters, and growth and differentiation factor can

activate kinases, which can occur as cell surface receptors or as the activator of the final effector

protein, but can also occur along the signal transduction pathway. Cascades of kinases occur, as well

as kinases sensitive to second messenger molecules. This system allows for the amplification of

20   weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many

weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the

extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such

as metabolic enzyme activity, proliferation; cell growth and differentiation, cell adhesion, and cell

cycle progression.

25   Protein phosphatases are characterized as either serine/threonine- or tyrosine-specific based

on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual

specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The

protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated

hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle

30   and cell signaling processes.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal

and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations,

such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle

contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic

35   metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated

into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Op. Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin dependent kinases, ERK family MAP kinases,

3

PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, supra).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppresant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, supra).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci 22:245-251). One of them, PP5, contains regulatory domains with tetratriticopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids in vitro and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue, which executes a nucleophilic attack on the phosphate moiety during catalysis. Receptor PTPs are made up of an N-terminal extracellular domain

4

of variable length, a transmembrane region, and a cytoplasmic region that generally contains two

copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the

second copy apparently affects the substrate specificity of the first. The extracellular domains of

some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains

5   (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains

(PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size

and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte

activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating

10 cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility,

most of them via the integrin/tyrosine kinase signalling pathway (reviewed in Neel, B.G. and N.K.

Tonks (1997) Curr. Op. Cell Bio. 9:193-204). CD45 phosphatases regulate signal transduction and

lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-32). Soluble

PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these

15 molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling

by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell

receptor and c-Kit (reviewed in Neel and Tonks, supra). M-phase inducer phosphatase plays a key

role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell

division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. 87:5139-5143). In addition, the genes

20 encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or

rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma,

leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992)

Annu. Rev. Cell Biol. 8:463-493). Many PTKs are encoded by oncogenes, and it is well known that

oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore

25 possible that PTPs may serve to prevent or reverse cell transformation and the growth of various

cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies

showing that overexpression of PTP can suppress transformation in cells and that specific inhibition

of PTP can enhance cell transformation (Charbonneau and Tonks, supra).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs.

30 DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are

primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C.

The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular

signal regulated kinase), JNK (c-Jun N terminal kinase), and p38 on both tyrosine and threonine

5

apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura. M. et al. (1999) J. Natl. Cancer Inst. 91:1820-8).

The discovery of new phosphatases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system disorders, neurological disorders, developmental disorders and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of phosphatases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, phosphatases, referred to collectively as "PP" and individually as "PP-1," "PP-2," "PP-3," "PP-4," "PP-5," "PP-6," "PP-7," "PP-8," and "PP-9." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-9.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-9. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:10-18.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90%

5   sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the

10   recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a

15   biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the

20   polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group

25   consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of

30   SEQ ID NO:10-18, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a naturally occurring polynucleotide sequence having at least 90% sequence

5   identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions

10 whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said

15 target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)

20 an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide

25 comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected

30 from the group consisting of SEQ ID NO:1-9, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining

of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:10-18, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii),

and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a

reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PP" refers to the amino acid sequences of substantially purified PP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

An "allelic variant" is an alternative form of the gene encoding PP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PP or a polypeptide with at least one functional characteristic of PP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,

12

hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include:

5   asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

10   molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known

15   in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

20   The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or

25   synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

30   immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PP or fragments of PP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein. i.e.. the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a

15

A "fragment" is a unique portion of PP or the polynucleotide encoding PP which is identical
in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the
entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a
fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment
5   used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,
15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid
residues in length. Fragments may be preferentially selected from certain regions of a molecule. For
example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected
from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain
10   defined sequence. Clearly these lengths are exemplary, and any length that is supported by the
specification, including the Sequence Listing, tables, and figures, may be encompassed by the present
embodiments.

A fragment of SEQ ID NO:10-18 comprises a region of unique polynucleotide sequence that
specifically identifies SEQ ID NO:10-18, for example, as distinct from any other sequence in the
15   genome from which the fragment was obtained. A fragment of SEQ ID NO:10-18 is useful, for
example, in hybridization and amplification technologies and in analogous methods that distinguish
SEQ ID NO:10-18 from related polynucleotide sequences. The precise length of a fragment of SEQ
ID NO:10-18 and the region of SEQ ID NO:10-18 to which the fragment corresponds are routinely
determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

20   A fragment of SEQ ID NO:1-9 is encoded by a fragment of SEQ ID NO:10-18. A fragment
of SEQ ID NO:1-9 comprises a region of unique amino acid sequence that specifically identifies SEQ
ID NO:1-9. For example, a fragment of SEQ ID NO:1-9 is useful as an immunogenic peptide for the
development of antibodies that specifically recognize SEQ ID NO:1-9. The precise length of a
fragment of SEQ ID NO:1-9 and the region of SEQ ID NO:1-9 to which the fragment corresponds are
25   routinely determinable by one of ordinary skill in the art based on the intended purpose for the
fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon
(e.g., methionine) followed by an open reading frame and a translation termination codon. A "full
length" polynucleotide sequence encodes a "full length" polypeptide sequence.

30   "Homology" refers to sequence similarity or, interchangeably, sequence identity, between two
or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to
the percentage of residue matches between at least two polynucleotide sequences aligned using a
standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in
35   the sequences being compared in order to optimize alignment between two sequences, and therefore

achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as

least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such

lengths are exemplary only, and it is understood that any fragment length supported by the sequences

shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which

percentage identity may be measured.

5   Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences

that all encode substantially the same protein.

        The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to

10 the percentage of residue matches between at least two polypeptide sequences aligned using a

standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment

methods take into account conservative amino acid substitutions. Such conservative substitutions,

explained in more detail above, generally preserve the charge and hydrophobicity at the site of

substitution, thus preserving the structure (and therefore function) of the polypeptide.

15   Percent identity between polypeptide sequences may be determined using the default parameters

of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment

program (described and referenced above). For pairwise alignments of polypeptide sequences using

CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and

"diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with

20 polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity"

between aligned polypeptide sequence pairs.

        Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise

comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12

(April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

25   *Matrix: BLOSUM62*

        Open Gap: 11 and Extension Gap: 1 penalties

        Gap x drop-off: 50

        Expect: 10

        Word Size: 3

30   *Filter: on*

        Percent identity may be measured over the length of an entire defined polypeptide sequence, for

example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for

example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance,

a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150

18

contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain
5   DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

10   "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency
15   of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive
20   annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic
25   strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

30   High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may

sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency

5   conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0 t$ or $R_0 t$ analysis) or formed between one nucleic acid

10   sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

15   "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PP which is capable

20   of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

25   The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PP.

30   The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

5   "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

10   "Post-translational modification" of an PP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PP.

           "Probe" refers to nucleic acid sequences encoding PP, their complements, or fragments
15 thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA
20 polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

           Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,
25 or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

           Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold
30 Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs

MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PP, nucleic acids encoding PP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,

23

or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting

polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human phosphatases (PP), the polynucleotides encoding PP, and the use of these compositions for the diagnosis, treatment, or prevention of immune system disorders, neurological disorders, developmental disorders and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns

Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of each polypeptide of the invention, and these properties establish that the claimed polypeptides are phosphatases. For example, SEQ ID NO:9 is 60% identical to Drosophila melanogaster MAP kinase phosphatase (GenBank ID g6714641) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.00E-104, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a dual-specificity phosphatase catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:9 is a MAP kinase. SEQ ID NO:1-8 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-9 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:10-18 or that distinguish between SEQ ID NO:10-18 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 1363431F6 is the identification number of an Incyte cDNA sequence, and LUNGNOT12 is the cDNA library from which

26

it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries. Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of

5   genomic DNA. For example, g2961443.v113.gs_2.nt is the identification number of a Genscan-predicted coding sequence, with g2961443 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching"

10   algorithm (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

15   Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

20   The invention also encompasses PP variants. A preferred PP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PP amino acid sequence, and which contains at least one functional or structural characteristic of PP.

         The invention also encompasses polynucleotides which encode PP. In a particular embodiment,

25   the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18, which encodes PP. The polynucleotide sequences of SEQ ID NO:10-18, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

30   The invention also encompasses a variant of a polynucleotide sequence encoding PP. In particular such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence

about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:10-18. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PP.

5   It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in

10 accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PP, and all such variations are to be considered as being specifically disclosed.

        Although nucleotide sequences which encode PP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PP or its

15 derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PP and its derivatives without altering the encoded amino acid sequences include the production of RNA

20 transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

        The invention also encompasses production of DNA sequences which encode PP and PP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using

25 reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PP or any fragment thereof.

        Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:10-18 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and

30 S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

        Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of

DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68° C to 72° C.

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PP may be cloned in recombinant DNA molecules that direct expression of PP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"

breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

5   maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PP

10   itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PP, or any

15   part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

20   The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PP, the nucleotide sequences encoding PP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a

25   suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences

30   encoding PP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous

and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies) Ligation of sequences encoding PP into the vector's multiple cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription,

dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PP are needed, e.g. for the production of antibodies, vectors which direct high level expression of PP may be used. For example, vectors containing the strong, inducible SP6 or T7
5   bacteriophage promoter may be used.

Yeast expression systems may be used for production of PP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into
10 the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PP. Transcription of sequences encoding PP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311).
15 Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.
20 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain
25 infective virus which expresses PP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
30 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

using expression vectors which may contain viral origins of replication and/or endogenous expression
elements and a selectable marker gene on the same or on a separate vector. Following the introduction
of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being
switched to selective media. The purpose of the selectable marker is to confer resistance to a selective

5   agent, and its presence allows growth and recovery of cells which successfully express the introduced
sequences. Resistant clones of stably transformed cells may be propagated using tissue culture
techniques appropriate to the cell type.

        Any number of selection systems may be used to recover transformed cell lines. These include,
but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase

10   genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232;
Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be
used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers
resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to
chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980)

15   Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.)
Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements
for metabolites. (See, e.g., Hartman, S.C. and R C. Mulligan (1988) Proc. Natl. Acad. Sci. USA
85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß
glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used.

20   These markers can be used not only to identify transformants, but also to quantify the amount of
transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A.
(1995) Methods Mol. Biol. 55:121-131.)

        Although the presence/absence of marker gene expression suggests that the gene of interest is
also present, the presence and expression of the gene may need to be confirmed. For example, if the

25   sequence encoding PP is inserted within a marker gene sequence, transformed cells containing
sequences encoding PP can be identified by the absence of marker gene function. Alternatively, a
marker gene can be placed in tandem with a sequence encoding PP under the control of a single
promoter. Expression of the marker gene in response to induction or selection usually indicates
expression of the tandem gene as well.

30   In general, host cells that contain the nucleic acid sequence encoding PP and that express PP
may be identified by a variety of procedures known to those of skill in the art. These procedures
include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and
protein bioassay or immunoassay techniques which include membrane, solution, or chip based
technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PP may be designed to contain signal sequences which direct secretion of PP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells

Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PP encoding sequence and the heterologous protein sequence, so that PP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PP. At least one and up to a plurality of test compounds may be screened for specific binding to PP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds

involves producing appropriate cells which express PP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, _Drosophila_, or E. coli. Cells expressing PP or cell membrane fractions which contain PP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PP or the compound is analyzed.

5   An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PP, either in solution or affixed to a solid support, and detecting the binding of PP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

10   Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

          PP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PP activity,

15   wherein PP is combined with at least one test compound, and the activity of PP in the presence of a test compound is compared with the activity of PP in the absence of the test compound. A change in the activity of PP in the presence of the test compound is indicative of a compound that modulates the activity of PP. Alternatively, a test compound is combined with an _in vitro_ or cell-free system comprising PP under conditions suitable for PP activity, and the assay is performed. In either of these

20   assays, a test compound which modulates the activity of PP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

          In another embodiment, polynucleotides encoding PP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES)

25   cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989)

30   Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D.

37

the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PP may also be manipulated in vitro in ES cells derived from human
5 blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PP can also be used to create "knockin" humanized animals (pigs) or
10 transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal
15 inbred to overexpress PP, e.g., by secreting PP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PP and phosphatases. In addition, the expression of PP is closely associated with
20 cell proliferation, cancer, and in particular with colon cancer. Therefore, PP appears to play a role in immune system disorders, neurological disorders, developmental disorders and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased PP expression or activity, it is desirable to decrease the expression or activity of PP. In the treatment of disorders associated with decreased PP expression or activity, it is desirable to increase the expression or
25 activity of PP.

Therefore, in one embodiment, PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common
30 variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke,

cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural

5   empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

10 nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

15 akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary

20 abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell

25 proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall

30 bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

       In another embodiment, a vector capable of expressing PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or

with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those listed above.

In a further embodiment, an antagonist of PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds PP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PP may be produced using methods which are generally known in the art. In particular, purified PP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PP. Antibodies to PP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions,

peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and <u>Corynebacterium parvum</u> are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at

5   least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PP may be prepared using any technique which provides for the

10 production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

15 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single

20 chain antibodies may be adapted, using methods known in the art, to produce PP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing <u>in vivo</u> production in the lymphocyte population

25 or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PP may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion

30 of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al.

specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two

5   non-interfering PP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PP-antibody complex divided by the molar

10   concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PP epitopes, represents the average affinity, or avidity, of the antibodies for PP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12}

15   L/mole are preferred for use in immunoassays in which the PP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical

20   Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PP-antibody

25   complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene

30   expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

5   Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other

10 systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in

15 the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

20 Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g.,

25 against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV), fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PP expression or regulation causes disease, the expression of PP

30 from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PP are

ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

5   Expression vectors that may be effective for the expression of PP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV),

10   SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the

15   FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver

20   polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

25   In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences required for

30   efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-

1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4+ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PP to cells which have one or more genetic abnormalities with respect to the expression of PP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PP to target cells which have one or more genetic abnormalities with respect to the expression of PP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the

163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

5   In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This

10   subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PP-coding RNAs and the synthesis of high levels of PP in vector transduced cells. While alphavirus infection is typically

15   associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PP into a variety of cell types. The specific transduction of a subset of cells in a population may require

20   the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can

25   be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A

30   complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PP

encoding PP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which

generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.
Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various
formulations are commonly known and are thoroughly discussed in the latest edition of Remington's
Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PP,

5   antibodies to PP, and mimetics, agonists, antagonists, or inhibitors of PP.

The compositions utilized in this invention may be administered by any number of routes
including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,
intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,
sublingual, or rectal means.

10   Compositions for pulmonary administration may be prepared in liquid or dry powder form.
These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case
of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting
formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins),
recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled

15   the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S.
Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle
injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active
ingredients are contained in an effective amount to achieve the intended purpose. The determination of

20   an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of
macromolecules comprising PP or fragments thereof. For example, liposome preparations containing a
cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the
macromolecule. Alternatively, PP or a fragment thereof may be joined to a short cationic N-terminal

25   portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into
the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999)
Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell
culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys,

30   or pigs. An animal model may also be used to determine the appropriate concentration range and route
of administration. Such information can then be used to determine useful doses and routes for
administration in humans.

symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can

5   be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of

10   administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time

15   and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and

20   methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

25   In another embodiment, antibodies which specifically bind PP may be used for the diagnosis of disorders characterized by expression of PP, or in assays to monitor patients being treated with PP or agonists, antagonists, or inhibitors of PP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PP include methods which utilize the antibody and a label to detect PP in human body fluids or in extracts of cells or tissues. The

30   antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PP expression. Normal or

standard values for PP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PP expressed in subject, control, and

5 disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and

10 quantify gene expression in biopsied tissues in which expression of PP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PP, and to monitor regulation of PP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PP or closely related molecules may be used to

15 identify nucleic acid sequences which encode PP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%

20 sequence identity to any of the PP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:10-18 or from genomic sequences including promoters, enhancers, and introns of the PP gene.

Means for producing specific hybridization probes for DNAs encoding PP include the cloning of polynucleotide sequences encoding PP or PP derivatives into vectors for the production of mRNA

25 probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ³²P or ³⁵S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

30 Polynucleotide sequences encoding PP may be used for the diagnosis of disorders associated with expression of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of

immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease,

5   dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob

10 disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular

15 dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy,

20 corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such

25 as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

30 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.   The polynucleotide

35 sequences encoding PP may be used in Southern or northern analysis, dot blot, or other

52

membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PP may be useful in assays that detect
5   the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the
10 presence of altered levels of nucleotide sequences encoding PP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PP, a
15 normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.
20 Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the
25 patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development
30 of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further

involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be

5   employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans.

10 Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR

15 products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which

20 assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

25 Methods which may also be used to quantify the expression of PP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of

30 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large

numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the

5   activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PP, fragments of PP, or antibodies specific for PP may be used as

10   elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

15   quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

20   hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the

25   case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

30   molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a

55

from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data

5   after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is

10   important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

      In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the

15   present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

      Another particular embodiment relates to the use of the polypeptide sequences of the present

20   invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

25   analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent

30   such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for

example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

5   A proteomic profile may also be generated using antibodies specific for PP to quantify the levels of PP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of

10   methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

         Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J.

15   Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

         In another embodiment, the toxicity of a test compound is assessed by treating a biological

20   sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

25   individual proteins and comparing these partial sequences to the polypeptides of the present invention.

         In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with

30   the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

         Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,

57

(1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

5   In another embodiment of the invention, nucleic acid sequences encoding PP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal

10   mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet.

15   7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

          Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map

20   data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

25   In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery

30   techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc.,

58

among normal, carrier, or affected individuals.

In another embodiment of the invention, PP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PP, or fragments thereof, and washed. Bound PP is then detected by methods well known in the art. Purified PP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PP specifically compete with a test compound for binding PP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PP.

In additional embodiments, the nucleotide sequences which encode PP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/177,719, U.S. Ser. No. 60/178,988, U.S. Ser. No. 60/184,959, and U.S. Ser. No. 60/190,142, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database

suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

5   Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA

10   purification kit (Ambion, Austin TX).

        In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units

15   5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs

20   were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX

25   DH10B from Life Technologies.

II.   Isolation of cDNA Clones

        Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC

30   Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

        Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-

well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using

PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner

5   (Labsystems Oy, Helsinki, Finland).

III.   Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

Sequencing reactions were processed using standard methods or high-throughput instrumentation

such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler

10   (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared

using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as

the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were

15   carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI

PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI

protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997,

supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques

20   disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector,

linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based

on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA

sequences or translations thereof were then queried against a selection of public databases such as the

25   GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS,

DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM.

(HMM is a probabilistic approach which analyzes consensus primary structures of gene families.

See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were

performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA

30   sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank

cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding

sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length.

The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt,

5   BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the

10   MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second

15   column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

20   The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:10-18. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV.   Identification and Editing of Coding Sequences from Genomic DNA

25   Putative phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an

30   assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for phosphatases. Potential phosphatases were also identified by homology to Incyte cDNA

sequences that had been annotated as phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V.   Assembly of Genomic Sequence Data with cDNA Sequence Data
"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence.

Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI.   Chromosomal Mapping of PP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:10-18 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:10-18 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII.   Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene

and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of

cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD

5   database (Incyte Genomics, Palo Alto CA).

VIII.   Extension of PP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was

10   synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

15   Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$,

20   and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2:

25   94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

30   Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For

shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose

gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were

5   religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs,

and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing

media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x

carb liquid media.

10   The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham

Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1:

94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4

repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN

reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified

15   using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2,

v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC

DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle

sequencing ready reaction kit (Applied Biosystems).

         In like manner, full length polynucleotide sequences are verified using the above procedure or

20   are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides

designed for such extension, and an appropriate genomic library.

IX.   **Labeling and Use of Individual Hybridization Probes**

         Hybridization probes derived from SEQ ID NO:10-18 are employed to screen cDNAs, genomic

DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is

25   specifically described, essentially the same procedure is used with larger nucleotide fragments.

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National

Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-³²P] adenosine

triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston

MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size

30   exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per

minute of the labeled probe is used in a typical membrane-based hybridization analysis of human

genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or

membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16
hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature
under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.
Hybridization patterns are visualized using autoradiography or an alternative imaging means and
5   compared.

X.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing
photolithography, piezoelectric printing (ink-jet printing. See, e.g., Baldeschweiler, supra.), mechanical
microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned
10   technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested
substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure
analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a
substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be
produced using available methods and machines well known to those of ordinary skill in the art and may
15   contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470;
Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol.
16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may
comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be
20   selected using software well known in the art such as LASERGENE software (DNASTAR). The array
elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the
biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.
After hybridization, nonhybridized nucleotides from the biological sample are removed, and a
fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser
25   desorbtion and mass spectrometry may be used for detection of hybridization. The degree of
complementarity and the relative abundance of each polynucleotide which hybridizes to an element on
the microarray may be assessed. In one embodiment, microarray preparation and usage is described in
detail below.

Tissue or Cell Sample Preparation

30   Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and
poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is
reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first
strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM
dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse

transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with

GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription

from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and

incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified

using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and

resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is

amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses

primers complementary to the vector sequences flanking the cDNA insert. Array elements are

amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5

μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia

Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water

washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C

oven.

Array elements are applied to the coated glass substrate using a procedure described in US

Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average

concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic

apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate

buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in

0.2% SDS and distilled water as before.

Hybridization

mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI.   Complementary Polynucleotides

Sequences complementary to the PP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PP-encoding transcript.

XII.   Expression of PP

Expression and purification of PP is achieved using bacterial or virus-based expression systems. For expression of PP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PP is synthesized as a fusion protein with, e.g., glutathione S-

enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using

5   commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PP obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX, where applicable.

10 **XIII.   Functional Assays**

PP function is assessed by expressing the sequences encoding PP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the

15 cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector.

20 Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA

25 content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the

30 cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human

immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PP Specific Antibodies

PP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PP activity by, for example, binding the peptide or PP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PP Using Specific Antibodies

Naturally occurring or recombinant PP is substantially purified by immunoaffinity chromatography using antibodies specific for PP. An immunoaffinity column is constructed by covalently coupling anti-PP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PP is collected.

XVI.   Identification of Molecules Which Interact with PP

e.g.. Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PP, washed, and any wells with labeled PP complex are assayed. Data obtained using different concentrations of PP are used to calculate values for the number, affinity, and association of PP with the candidate molecules.

5   Alternatively, molecules interacting with PP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between 10 the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PP Activity

PP activity is measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). PP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 15 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH and the increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of PP in the assay (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62).

In the alternative, PP activity is determined by measuring the amount of phosphate removed 20 from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM enzyme in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% 2-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then 25 centrifuged at 12,000 × g for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XVIII. Identification of PP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVI. PP 30 activity is measured for each well and the ability of each compound to inhibit PP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PP activity.

XIX.   Identification of PP Substrates

A PP "substrate-trapping" assay takes advantage of the increased substrate affinity that may 35 be conferred by certain mutations in the PTP signature sequence. PP bearing these mutations form a

stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of PP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of PP mutants in Escherichia coli, DNA

5   fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding PP or a glutathione S-transferase (GST)-PP fusion protein. PP mutants are expressed in E. coli and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 µg of CsCl-purified DNA per 10-cm dish of cells or 8 µg per 6-cm dish. Forty-

10   eight hours after transfection, cells are stimulated with 100ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 µg/ml leupeptin/5 µg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). PP is immunoprecipitated from lysates with an

15   appropriate antibody. GST-PP fusion proteins are precipitated with glutathione-Sepharose, 4 µg of mAb or 10 µl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint. A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

20   Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious

25   to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1269556	1	1269556CD1	10	1269556CB1
1849177	2	1849177CD1	11	1849177CB1
3402521	3	3402521CD1	12	3402521CB1
1723447	4	1723447CD1	13	1723447CB1
1878677	5	1878677CD1	14	1878677CB1
6205333	6	6205333CD1	15	6205333CB1
7472032	7	7472032CD1	16	7472032CB1
2522707	8	2522707CD1	17	2522707CB1
637331	9	637331CD1	18	637331CB1

Table 2

Polypeptide ID NO:	Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
		g2352922	1.8e-53	glucose-6-phosphatase [Haplochromis nubilus]
		g1399103	1.5e-92	phosphatidylinositol (4,5) bisphosphate 5-phosphatase homolog [Homo sapiens]
		g3880627	9.9e-60	similar to Protein phosphatase 2C (2 domains) [Caenorhabditis elegans]
		g2804429	1.5e-64	similar to the protein phosphatase 2C family [Caenorhabditis elegans]
		g6164723	8.9e-157	F-box protein Fbl4 [Homo sapiens]
		g4104681	6.7e-49	MKP-1 like protein tyrosine phosphatase [Homo sapiens]
		g1777755	4.8e-46	protein tyrosine phosphatase PTPCAAX1 [Homo sapiens]
		g1065967	8.2e-29	similar to other protein phosphatases 1, 2A and 2B [Caenorhabditis elegans]
		g6714641	1.00E-104	MAP kinase phosphatase [Drosophila melanogaster]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1269556CD1	395	S100 T360 S59 T187		transmembrane domain V25-F41, I162-P188	HMMER MOTIFS
					GLUCOSE-6-PHOSPHATASE G6PASE GLYCOGEN PD014226: G8-L330	BLAST-PRODOM
2	1949277CD1	189	S62 S167 T11 T31 S97 T10 S77 T90 T99	N180	Spscan:M1-A47	SPSCAN MOTIFS
					Inositol polyphosphate phosphatase family, c:M1-R143	HMMER-PFAM
					Inositol polyphosphate p PF00783A: F17-L26 R94 L103	BLIMPS_PFAM
					PROTEIN INOSITOL HYDROLASE 5-PHOSPHATASE: PD002029:D14-R107	BLAST-PRODOM
3	3402521CD1	304	S148 T252 S179 S20 S122 T208 T220 S244 S255		signal_cleavage: M1-T22	SPSCAN
					Protein phosphatase 2C PP2C:Q128-R172	HMMER-PFAM
					Phosphate aminotransferase, PF00040:V111-I118	BLIMPS-PRODOM
4	1723447CD1	440	S50 T257 S306 T425 S431 T278 T364		Protein phosphatase 2C PP2C: L187 L265 Q326-L403	HMMER-PFAM
					PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE FAMILY PP2C ISOFORM: PD001101:G322-L403	BLAST-PRODOM
					PROTEIN PHOSPHATASE 2C DM00377|P49596|1-295: A191-I262	BLAST-DOMO
					Protein Phosphatase 2C BL01032:Y120-G129 L187-G204 R328-D341 D376-D388	BLIMPS-BLOCKS

Table 3 (cont.)

Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1573677CD1	303	T34 S177 S235 S288 S267 T27 S87 T102 S212 T218 S239	N184 N210		MOTIFS
2853333CD1	157	S136 T59 S126	N51	Dual specificity phosphatase, catalytic domain: DSPc: G19-Q157	HMMER PFAM
				VH1-TYPE DUAL SPECIFICITY PHOSPHATASE: DM03823|P28562|169-314: G19-L158; DM08829|P40479|48-197: M48-K161	BLAST DOMO
				Tyrosine Phosphatase: L102-L114	MOTIFS
				Tyrosine specific protein phosphatases active site: D81-L184	PROFILESCAN BLIMPS-BLOCKS BLIMPS-PRINTS
				Tyrosine specific protein, BL00383E: V111-P121	BLIMPS-BLOCKS
				Tyrosine specific protein phosphatases active site: K100-V139	SPSCAN
				PROTEIN TYROSINE PHOSPHATASE, PR00700D:C108-L126	BLIMPS-PRINTS
3033838CD1	177	S35 S163 S23 T49 T66		SIMILAR TO OTHER PROTEIN PHOSPHATASES 1: PD140725|075-G268	BLAST PRODOM
3037567CD1	467	T69 T72 T94 S132 S165 T186 S234 S306 S363 Y138	N133		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	637331CD1	484	S21 S25 S36 S129 S207 S219 T245 T271 T277 S324 T43 T59 S63 S231 T250 T374 S419 T452 Y428	N55 N102 N131 N269 N322 N339 N372 N448	Dual specificity phosphatase, catalytic domain: S307:C436	HMMER_PFAM
					VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823 P28562|169-314:P309-A418 A56115|51-336:L302-A418	BLAST_DOMO

Table 4

Nucleotide ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
	1363......	1569	1-26, 597-715, 106-152	1363431F6 (LUNGNOT12)	409	860
				885874R1 (PANCNOT05)	114	736
				2107745H1 (BRAITUT03)	1	252
				1667068T6 (BMARNOT03)	799	1551
				1808801F6 (PROSTUT12)	1317	1569
				1269556T1 (BRAINOT09)	904	1552
	1269......	1158	1-31	1267666F1 (BRAINOT09)	858	1158
				2441905F6 (EOSITXT01)	1	569
				1854682F6 (HNT3AZT01)	500	1136
	3402521......	1110	564-907	6985930H1 (BRAIFER05)	113	657
				3402521H1 (ESOGNOT03)	590	844
				6570748H1 (MCLDTXN05)	806	1110
				g2961443.v113.gs_2.nt	1	945
	6932......	1837	1-74, 1418-1837	6932732H1 (BRSTNON02)	1230	1837
				3277266H1 (PROSBPT06)	264	508
				g4309923.v113.gs_1.nt.edit	1	795
	1878......	1504	1-442, 1314-1504	675272SJ1 (SINTFER02)	65	225
				2113407T6 (BRAITUT03)	584	1270
				1628276F6 (COLNPOT01)	447	1008
				1878677F6 (LEUKNOT03)	1230	1504
				2605756H1 (LUNGTUT07)	1	264
				5096112H1 (EPIMNON05)	1033	1307
				6486613H1 (MIXDUNB01)	582	1090
	6205......	1407	347-428, 1255-1407	2770081F6 (COLANOT02)	111	592
				3334286F6 (BRAIFET01)	861	1407
				2953210F6 (CONNNCT02)	1	405
				6311439H1 (NERDTDN03)	338	865
	7472......	549		6205333H1 (PITUNON01)	463	890
				g4156143.v113.gs_2.nt.edit	1	549

Table 4 (cont.)

17	252?707CB1	1587	1-756	2?5553?H1 (OVART?T01)	80?	1026
				2493276T6 (ADRETUT05)	1006	1575
				496386R6 (HNT2NOT01)	1	548
				1342043F6 (COLNTUT03)	123?	1587
				2522707F7 (BRA?TUT21)	447	968
18	63733?CB1	1771	1279 1339, 581 1070	2556630T6 (OVARTUT01)	129?	1771
				387789H1 (THYMNOT02)	35	303
				70905944V1	100?	1629
				70905746V1	335	1003
				70909374V1	913	1570
				g3894502	1	432
				2818170F6 (BRSTNOT14)	1	239

Table 5

nucleotide SEQ ID NO:	Incyte Project ID	Representative Library
	1269556CB1	BRAINOTC9
	1849177CB1	LUNGFETC3
	3402521CB1	BRAIFEPC5
	1723447CB1	BRAITUTC3
	1878677CB1	BRSTNOT12
	6205333CB1	KIDNTUT16
	2522707CB1	COLNTUT03
	637331CB1	NEJTGMT01

Table 6

Library	Vector	Library Description
BRAINOT09	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
LUNGFET03	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAITUT03	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
BRSTNOT12	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 32-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included benign hypertension and atherosclerotic coronary artery disease.
KIDNTUT16	pINCY	Library was constructed using RNA isolated from left pole kidney tumor tissue removed from a 53-year-old Caucasian female during a nephroureterectomy. Pathology indicated grade 2 renal cell carcinoma. Patient history included hyperlipidemia, cardiac dysrhythmia, metrorrhagia, cerebrovascular disease, atherosclerotic coronary artery disease, and tobacco abuse. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
COLNTUT03	pINCY	Library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma. One lymph node contained metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer, and prostate cancer.
NEUTGMT01	PSPORT1	Library was constructed using 1 microgram of polyA RNA isolated from peripheral blood granulocytes collected by density gradient centrifugation through Ficoll-Hypaque. The cells were isolated from buffy coat units obtained from 20 unrelated male and female donors. Cells were cultured in 10 nM GM-CSF for 1 hour before washing and harvesting for total RNA preparation.

Table 7

	Description	Reference	Parameter Threshold
	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
mbler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs*: Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448. Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489).	*ESTs*: fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences*: fastx score=100 or greater
	A Blocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits*: Probability value= 1.0E-3 or less *Signal peptide hits*: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park. CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

5   a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

          b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

          c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. and

10   d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-9.

15 3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID

20 NO:10-18.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

25 7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1. the method comprising:

30   a) culturing a cell under conditions suitable for expression of the polypeptide. wherein said cell is transformed with a recombinant polynucleotide. and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1. and

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

5   a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

10   e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15   13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization

20   complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

30   b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

18. A method for treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

    a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

    b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

    a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

    b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

    a) combining the polypeptide of claim 1 with at least one test compound under suitable

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

5   a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

     b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

     c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in

10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

15 comprising:

     a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

     b) detecting altered expression of the target polynucleotide, and

     c) comparing the expression of the target polynucleotide in the presence of varying amounts of

20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

     a) treating a biological sample containing nucleic acids with the test compound;

     b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at

25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

     c) quantifying the amount of hybridization complex; and

30   d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
      BANDMAN, Olga
      MATHUR, Preete
      TANG, Y. Tom
      AZIMZAI, Yalda
      YUE, Henry
      BAUGHN, Mariah R.
      HILLMAN, Jennifer L.
      LAL, Preeti
      WANG, Eureka
      GANDHI, Ameena R.
      POLICKY, Jennifer L.

<120> PHOSPHATASES

<130> PI-0018 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/177,719; 60/178,988; 60/184,959; 60/190,142
<151> 2000-01-21; 2000-01-28; 2000-02-25; 2000-03-17

<160> 16
<170> PERL Program

<210> 1
<211> 385
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1269556CD1

<400> 1
Met Glu Ser Thr Leu Gly Ala Gly Ile Val Ile Ala Glu Ala Leu
 1   5   10   15
Gln Asn Gln Leu Ala Trp Leu Glu Asn Val Trp Leu Trp Ile Thr
                20   25   30
Phe Leu Gly Asp Pro Lys Ile Leu Phe Leu Phe Tyr Phe Pro Ala
                35   40   45
Ala Tyr Tyr Ala Ser Arg Arg Val Gly Ile Ala Val Leu Trp Ile
                50   55   60
Ser Leu Ile Thr Glu Trp Leu Asn Leu Ile Phe Lys Trp Phe Leu
                65   70   75
Phe Gly Asp Arg Pro Phe Trp Trp Val His Glu Ser Gly Tyr Tyr
                80   85   90
Ser Gln Ala Pro Ala Gln Val His Gln Phe Pro Ser Ser Cys Glu
                95   100   105
Thr Gly Pro Gly Ser Pro Ser Gly His Cys Met Ile Thr Gly Ala
                11.   ...   ...
Ala Leu Trp Pro Leu Met Thr Ala Leu Pro Ser Th. Val Ala Thr
                12.   ...
Arg Ala Arg Ser Arg Trp Val Arg Val Met Pro Ser Leu Ala Tyr
                13.   145   15.
Cys Thr Phe Leu Leu Ala Val Gly Leu Ser Arg Ile Phe Ile Leu
                155   160   165
Ala His Phe Pro His Gln Val Leu Ala Gly Leu Ile Thr Gly Ala
                17.   175   180
Val Leu Gly Trp Leu Met Thr Pro Arg Val Pro Met Glu Arg Glu
                1..   1...   ...
Leu Ser Phe Tyr Gly Leu Thr Ala Leu Ala Leu Met Leu Gly Thr
                ...   .

```
                      190                    195                    200
Ile His Val Asp Ser Arg Pro Phe Ala Ser Leu Ser Arg Asp Ser
                      245                    250                    255
Gly Ala Ala Leu Gly Leu Gly Ile Ala Leu His Ser Pro Cys Tyr
                      260                    265                    270
Ala Gln Val Arg Arg Ala Gln Leu Gly Asn Gly Gln Lys Ile Ala
                      275                    280                    285
Cys Leu Val Leu Ala Met Gly Leu Leu Gly Pro Leu Asp Trp Leu
                      290                    295                    300
Gly His Pro Pro Gln Met Gln Pro Leu Leu His Phe Gln Phe Pro
                      305                    310                    315
Gln Val His Pro Leu Ala Met Pro Ser Pro Gly Pro Arg Ala Leu
                      320                    325                    330
Gly Ser Ala His Val Gln Cys Pro Gly Ser Thr Ala His Pro Leu
                      335                    340                    345
Phe Leu Thr Ser Cys Val Pro Pro Phe Pro Phe Pro Pro Thr Lys
                      350                    355                    360
Pro Thr Leu Cys Asp His His Thr Pro Gly Gly Ser Pro Ile Pro
                      365                    370                    375
Phe Gln Pro Leu Ser Arg Pro Ser Pro Pro
                      380                    385
```

<210> 2
<211> 188
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1849177CD1

<400> 2
```
Met Ala His Cys Ala Ala Ala Asp Val Thr Thr Arg Phe Asp Glu
 1                    5                     10                    15
Val Phe Trp Phe Gly Asp Phe Asn Phe Arg Leu Ser Gly Gly Arg
                     20                    25                     30
Thr Val Val Asp Ala Leu Leu Cys Gln Gly Leu Val Val Asp Val
                     35                    40                     45
Pro Ala Leu Leu Gln His Asp Gln Leu Ile Arg Glu Met Arg Lys
                     50                    55                     60
Gly Ser Ile Phe Lys Gly Phe Gln Glu Pro Asp Ile His Phe Leu
                     65                    70                     75
Pro Ser Tyr Lys Phe Asp Ile Gly Lys Asp Thr Tyr Asp Ser Thr
                     80                    85                     90
Ser Lys Gln Arg Thr Pro Ser Tyr Thr Asp Arg Val Leu Tyr Arg
                     95                    100                   105
Ser Arg His Lys Gly Asp Ile Cys Pro Val Ser Tyr Ser Ser Cys
                     110                   115                   120
Pro Gly Ile Lys Thr Ser Asp His Arg Pro Val Tyr Gly Leu Phe
                     125                   130                   135
Arg Val Lys Val Arg Pro Gly Arg Asp Asn Ile Pro Leu Ala Ala
                     140                   145                   150
Gly Lys Phe Asp Arg Glu Leu Tyr Leu Leu Gly Ile Lys Arg Arg
                     155                   160                   165
Ile Ser Lys Glu Ile Gln Arg Gln Gln Ala Leu Gln Ser Gln Asn
                     170                   175                   180
Ser Ser Thr Ile Cys Ser Val Ser
                     185
```

<210> 3
<211> 304
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3402521CD1

```
<400> 3
Met Phe Ser Val Leu Ser Tyr Gly Arg Leu Val Ala Arg Ala Val
1               5                   10                  15
Leu Gly Gly Leu Ser Gln Thr Asp Pro Arg Ala Gly Gly Gly Gly
                20                  25                  30
Gly Gly Asp Tyr Gly Leu Val Thr Ala Gly Cys Gly Phe Gly Lys
                35                  40                  45
Asp Phe Arg Lys Gly Leu Leu Lys Lys Gly Ala Cys Tyr Gly Asp
                50                  55                  60
Asp Ala Cys Phe Val Ala Arg His Arg Ser Ala Asp Val Leu Gly
                65                  70                  75
Val Ala Asp Gly Val Gly Gly Trp Arg Asp Tyr Gly Val Asp Pro
                80                  85                  90
Ser Gln Phe Ser Gly Thr Leu Met Arg Thr Cys Glu Arg Leu Val
                95                  100                 105
Lys Glu Gly Arg Phe Val Pro Ser Asn Pro Ile Gly Ile Leu Thr
                110                 115                 120
Thr Ser Tyr Cys Glu Leu Leu Gln Asn Lys Val Pro Leu Leu Gly
                125                 130                 135
Ser Ser Thr Ala Cys Ile Val Val Leu Asp Arg Thr Ser His Arg
                140                 145                 150
Leu His Thr Ala Asn Leu Gly Asp Ser Gly Phe Leu Val Val Arg
                155                 160                 165
Gly Gly Glu Val Val His Arg Ser Asp Glu Gln Gln His Tyr Phe
                170                 175                 180
Asn Thr Pro Phe Gln Leu Ser Ile Ala Pro Pro Glu Ala Glu Gly
                185                 190                 195
Val Val Leu Ser Asp Ser Pro Asp Ala Ala Asp Ser Thr Ser Phe
                200                 205                 210
Asp Val Gln Leu Gly Asp Ile Ile Leu Thr Ala Thr Asp Gly Leu
                215                 220                 225
Phe Asp Asn Met Pro Asp Tyr Met Ile Leu Gln Glu Leu Lys Lys
                230                 235                 240
Leu Lys Asn Ser Asn Tyr Glu Ser Ile Gln Gln Thr Ala Arg Ser
                245                 250                 255
Ile Ala Glu Gln Ala His Glu Leu Ala Tyr Asp Pro Asn Tyr Met
                260                 265                 270
Ser Pro Phe Ala Gln Phe Ala Cys Asp Asn Gly Leu Asn Val Arg
                275                 280                 285
Gly Gly Lys Pro Asp Asp Ile Thr Val Leu Leu Ser Ile Val Ala
                290                 295                 300
Glu Tyr Thr Asp


<210> 4
<211> 440
<212> PRT
<213> Homo sapiens

<220>
... misc_feature
...

<400> 4
Met Ser Ala Gly Trp Phe Arg Ala Arg ... Leu Pro Gly ...
1               .                   .
Leu Pro Ala Pro Arg Pro Pro Gly Pro His Ala Ser Pro Val Pro
                20                  .                   30
Tyr Arg Arg Pro Arg Phe Leu Arg Gly Ser Ser Ser Ser Pro Gly
                35                  40                  45
Ala Ala Asp Ala Ser Arg Arg Pro Asp Ser Arg Pro Val Arg Ser
                50                  .                   .
Pro Ala Arg Gly Arg Thr Leu Pro Trp Asn Ala Gly Tyr Ala Glu
                .                   .                   .
```

```
                     95                        100                        105
Glu Trp Leu Thr Leu Cys Pro Glu Glu Phe Leu Thr Gly His Tyr
                    110                        115                        120
Trp Ala Leu Phe Asp Gly His Gly Gly Pro Ala Ala Ala Ile Leu
                    121                        130                        135
Ala Ala Asn Thr Leu His Ser Cys Leu Arg Arg Gln Leu Glu Ala
                    140                        145                        150
Val Val Glu Gly Leu Val Ala Thr Gln Pro Pro Met His Leu Asn
                    155                        160                        165
Gly Arg Cys Ile Cys Pro Ser Asp Pro Gln Phe Val Glu Glu Lys
                    170                        175                        180
Gly Ile Arg Ala Glu Asp Leu Val Ile Gly Ala Leu Glu Ser Ala
                    181                        190                        195
Phe Gln Glu Cys Asp Glu Val Ile Gly Arg Glu Leu Glu Ala Ser
                    200                        205                        210
Gly Gln Met Gly Gly Cys Thr Ala Leu Val Ala Val Ser Leu Gln
                    211                        220                        225
Gly Lys Leu Tyr Met Ala Asn Ala Gly Asp Ser Arg Ala Ile Leu
                    230                        235                        240
Val Arg Arg Asp Glu Ile Arg Pro Leu Ser Phe Glu Phe Thr Pro
                    241                        250                        255
Glu Thr Glu Arg Gln Arg Ile Gln Gln Leu Ala Phe Val Tyr Pro
                    260                        265                        270
Glu Leu Leu Ala Gly Glu Phe Thr Arg Leu Glu Phe Pro Arg Arg
                    275                        280                        285
Leu Lys Gly Asp Asp Leu Gly Gln Lys Val Leu Phe Arg Asp His
                    290                        295                        300
His Met Ser Gly Trp Ser Tyr Lys Arg Val Glu Lys Ser Asp Leu
                    305                        310                        315
Lys Tyr Pro Leu Ile His Gly Gln Gly Arg Gln Ala Arg Leu Leu
                    316                        320                        325
Gly Thr Leu Ala Val Ser Arg Gly Leu Gly Asp His Gln Leu Arg
                    330                        340                        345
Val Leu Asp Thr Asn Ile Gln Leu Lys Pro Phe Leu Leu Ser Val
                    350                        360                        365
Pro Gln Val Thr Val Leu Asp Val Asp Gln Leu Glu Leu Gln Glu
                    366                        370                        375
Asp Asp Val Val Val Met Ala Thr Asp Gly Leu Trp Asp Val Leu
                    380                        385                        390
Ser Asn Glu Gln Val Ala Trp Leu Val Arg Ser Phe Leu Pro Gly
                    395                        400                        405
Asn Gln Glu Glu Pro Thr Gln Val Leu Lys Ala Gly Pro Asp Ala
                    410                        415                        420
Asp Thr Gln His Thr Gly Lys Gly Arg Gln Ser His Arg Gly Arg
                    425                        430                        435
Ala Gly Val Leu Arg
                    440
```

<210> 5
<211> 300
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1878677CD1

<400> 5
```
Met Ser Pro Val Phe Pro Met Leu Thr Val Leu Thr Met Phe Tyr
 1                   5                        10                         15
Tyr Ile Cys Leu Arg Arg Arg Ala Arg Thr Ala Thr Arg Gly Glu
                    20                        25                         30
Met Met Asn Thr His Arg Ala Ile Glu Ser Asn Ser Gln Thr Ser
                    35                        40                         45
Pro Leu Asn Ala Glu Val Val Gln Tyr Ala Lys Glu Val Val Asp
                    50                        55                         60
Phe Ser Ser His Tyr Gly Ser Glu Asn Ser Met Ser Tyr Thr Met
```

4/13

```
                    65                      70                      75
Trp Asn Leu Ala Gly Val Pro Asn Val Phe Pro Ser Ser Gly Asp
                    80                      85                      90
Phe Thr Gln Thr Ala Val Phe Arg Thr Tyr Gly Thr Trp Trp Asp
                    95                     100                     105
Gln Cys Pro Ser Ala Ser Leu Pro Phe Lys Arg Thr Pro Pro Asn
                   110                     115                     120
Phe Gln Ser Gln Asp Tyr Val Glu Leu Thr Phe Glu Gln Gln Val
                   125                     130                     135
Tyr Pro Thr Ala Val His Val Leu Glu Thr Tyr His Pro Gly Ala
                   140                     145                     150
Val Ile Arg Ile Leu Ala Cys Ser Ala Asn Pro Tyr Ser Pro Asn
                   155                     160                     165
Pro Pro Ala Glu Val Arg Trp Glu Ile Leu Trp Ser Glu Arg Pro
                   170                     175                     180
Thr Lys Val Asn Ala Ser Gln Ala Arg Gln Phe Lys Pro Cys Ile
                   185                     190                     195
Lys Gln Ile Asn Phe Pro Thr Asn Leu Ile Arg Leu Glu Val Asn
                   200                     205                     210
Ser Ser Leu Leu Glu Tyr Tyr Thr Glu Leu Asp Ala Val Val Leu
                   215                     220                     225
His Gly Val Lys Asp Lys Pro Val Leu Ser Leu Lys Thr Ser Leu
                   230                     235                     240
Ile Asp Met Asn Asn Ile Glu Asp Asp Ala Tyr Ala Glu Lys Asp
                   245                     250                     255
Gly Cys Gly Met Asp Ser Leu Asn Lys Lys Phe Ser Ser Ala Val
                   260                     265                     270
Leu Gly Glu Gly Pro Asn Asn Gly Tyr Phe Asp Lys Leu Pro Tyr
                   275                     280                     285
Glu Val Ser Gln Lys Tyr Leu Ala Ala Val Leu Asp Ile Thr Leu
                   290                     295                     300
```

```
<210> 6
<211> 188
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6205333CD1

<400> 6
Met Thr Ala Pro Ser Cys Ala Phe Pro Val Gln Phe Arg Gln Pro
  1               5                      10                      15
Ser Val Ser Gly Leu Ser Gln Ile Thr Lys Ser Leu Tyr Ile Ser
                   20                      25                      30
Asn Gly Val Ala Ala Asn Asn Lys Leu Met Leu Ser Ser Asn Gln
                   35                      40                      45
Ile Thr Met Val Ile Asn Val Ser Val Glu Val Val Asn Thr Leu
                   50                      55                      60
Tyr Glu Arg Ile Gln Tyr Met Gln Val Ile Thr Ala Arg Ser Pro
                   65                      70                      75
Lys Ser Arg Ser Tyr Ala Ile Ile Ser Ile Ile Ala Asn His Ile
                   80                      85                      90
His Ser Val Glu Met Lys Gln Gly Arg Thr Ile Leu Leu His Tyr Ala
                   95                     100                     105
Ala Gly Val Ser Arg Ser Ala Ala Leu Cys Leu Ala Tyr Leu Met
                   110                     115                     120
Lys Tyr His Ala Met Ser Leu Leu Asp Ala His Thr Trp Thr Lys
                   125                     130                     135
Ser Tyr Arg Ile Ile Ile Arg Ser Asn Ser Gly Phe Trp Glu Gln
                   140                     145                     150
```

Glu Val Arg Leu Met Ile Pro Leu
                185

<210> 7
<211> 182
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472032CD1

<400> 7
Met Gln Lys Arg Lys Pro Leu Ile Lys Leu Ser Asp Leu Leu Arg
1   5   10   15
Leu Ile Arg Tyr Tyr Glu Asn Ser Met Gly Glu Thr Ala Pro Met
                20   25   30
Ile Gln Ile Ile Ser His Arg Val Pro Leu Thr Thr Leu Val Asn
                35   40   45
His Gly Ser Thr Ile Gln Asp Glu Ile Trp Ile Arg Val Cys Lys
                50   55   60
Ala Thr Tyr Asp Thr Thr Leu Leu Glu Lys Glu Gly Thr His Val
                65   70   75
Leu Asp Arg Pro Phe Asp Asp Gly Ala Pro Pro Ser Asn Gln Ile
                80   85   90
Asp Asp Trp Leu Ser Leu Val Lys Ile Lys Phe His Lys Gln Pro
                95   100   105
His Cys Cys Thr Ala Val His Cys Ile Ala Gly Leu Gly Arg Ala
                110   115   120
Pro Val Leu Val Ala Leu Ala Leu Ile Glu Gly Lys Met Lys Asp
                125   130   135
Asp Asp Ala Val Gln Phe Ala Arg Lys Lys Gln Tyr Gly Ala Phe
                140   145   150
Asn Ser Lys Leu Thr Phe Val Phe Gly Glu Val Ser Ser Leu Lys
                155   160   165
Met Trp Leu His Phe Lys Asp Ser Ser Gly His Arg Tyr Asn Cys
                170   175   180
Cys Ser

<210> 8
<211> 402
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2522707CD1

<400> 8
Met Gln His Ile Val Gly Val Pro His Val Leu Val Arg Arg Gly
1   5   10   15
Leu Leu Gly Arg Asp Leu Phe Met Thr Arg Thr Leu Cys Ser Pro
                20   25   30
Gly Pro Ser Gln Pro Gly Glu Lys Arg Pro Glu Glu Val Ala Leu
                35   40   45
Gly Leu His His Arg Leu Pro Ala Leu Gly Arg Ala Leu Gly His
                50   55   60
Ser Ile Gln Gln Arg Ala Thr Ser Thr Ala Lys Thr Trp Trp Asp
                65   70   75
Arg Tyr Glu Glu Phe Val Gly Leu Asn Glu Val Arg Glu Ala Gln
                80   85   90
Gly Lys Val Thr Glu Ala Glu Lys Val Phe Met Val Ala Arg Gly
                95   100   105
Leu Val Arg Glu Ala Arg Glu Asp Leu Glu Val His Gln Ala Lys
                110   115   120
Leu Lys Glu Val Arg Asp Arg Leu Asp Arg Val Ser Arg Glu Asp

```
                        125                        130                        135
Ser Gln Tyr Leu Glu Leu Ala Thr Leu Glu His Arg Met Leu Gln
                        140                        145                        150
Glu Glu Lys Arg Leu Arg Thr Ala Tyr Leu Arg Ala Glu Asp Ser
                        155                        160                        165
Glu Arg Glu Lys Phe Ser Leu Phe Ser Ala Ala Val Arg Glu Ser
                        170                        175                        180
His Glu Lys Glu Arg Thr Arg Ala Glu Arg Thr Lys Asn Trp Ser
                        185                        190                        195
Leu Ile Gly Ser Val Leu Gly Ala Leu Ile Gly Val Ala Gly Ser
                        200                        205                        210
Thr Tyr Val Asn Arg Val Arg Leu Gln Glu Leu Lys Ala Leu Leu
                        215                        220                        225
Leu Glu Ala Gln Lys Gly Pro Val Ser Leu Gln Glu Ala Ile Arg
                        230                        235                        240
Glu Gln Ala Ser Ser Tyr Ser Arg Gln Gln Arg Asp Leu His Asn
                        245                        250                        255
Leu Met Val Asp Leu Arg Gly Leu Val His Ala Ala Gly Pro Gly
                        260                        265                        270
Gln Asp Ser Gly Ser Gln Ala Gly Ser Pro Pro Thr Arg Asp Arg
                        275                        280                        285
Asp Val Asp Val Leu Ser Ala Ala Leu Lys Glu Gln Leu Ser His
                        290                        295                        300
Ser Arg Gln Val His Ser Cys Leu Glu Gly Leu Arg Glu Gln Leu
                        305                        310                        315
Asp Gly Leu Glu Lys Thr Cys Ser Gln Met Ala Gly Val Val Gln
                        320                        325                        330
Leu Val Lys Ser Ala Ala His Pro Gly Leu Val Glu Pro Ala Asp
                        335                        340                        345
Gly Ala Met Pro Ser Phe Leu Leu Glu Gln Gly Ser Met Ile Leu
                        350                        355                        360
Ala Leu Ser Asp Thr Glu Gln Arg Leu Glu Ala Gln Val Asn Arg
                        365                        370                        375
Asn Thr Ile Tyr Ser Thr Leu Val Thr Cys Val Thr Phe Val Ala
                        380                        385                        390
Thr Leu Pro Val Leu Tyr Met Leu Phe Lys Ala Ser
                        395                        400
```

<210> 9
<211> 484
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6376310D1

<400> 9
```
Met Ala Leu Val Thr Val Gln Arg Ser Pro Thr Pro Ser Thr Thr
 1                    5                         10                         15
                                                                      
                                                                      
                                                                      
                                                                      
                                                                      
                                                                      
```

                    140   145   150
His Leu Asp Gly Asp Gly Gly Phe Ser Val Ser Thr Asp Asn Arg
                        155   160   165
Val His Ile Phe Lys Pro Val Ser Val Gln Ala Met Trp Ser Ala
                170   175   180
Leu Gln Ser Leu His Lys Ala Cys Glu Val Ala Arg Ala His Asn
                        185   190   195
Tyr Tyr Pro Gly Ser Leu Phe Leu Thr Trp Val Ser Tyr Tyr Glu
                200   205   210
Ser His Ile Asn Ser Asp Gln Ser Ser Val Asn Glu Trp Asn Ala
                        215   220   225
Met Gln Asp Val Gln Ser His Arg Pro Asp Ser Pro Ala Leu Phe
                230   235   240
Thr Asp Ile Pro Thr Glu Arg Glu Arg Thr Glu Arg Leu Ile Lys
                        245   250   255
Thr Lys Leu Arg Glu Ile Met Met Gln Lys Asp Leu Glu Asn Ile
                260   265   270
Thr Ser Lys Glu Ile Arg Thr Glu Leu Glu Met Gln Met Val Cys
                        275   280   285
Asn Leu Arg Glu Phe Lys Glu Phe Ile Asp Asn Glu Met Ile Val
                290   295   300
Ile Leu Gly Gln Met Asp Ser Pro Thr Gln Ile Phe Glu His Val
                        305   310   315
Phe Leu Gly Ser Glu Trp Asn Ala Ser Asn Leu Glu Asp Leu Gln
                320   325   330
Asn Arg Gly Val Arg Tyr Ile Leu Asn Val Thr Arg Glu Ile Asp
                        335   340   345
Asn Phe Phe Pro Gly Val Phe Glu Tyr His Asn Ile Arg Val Tyr
                350   355   360
Asp Glu Glu Ala Thr Asp Leu Leu Ala Tyr Trp Asn Asp Thr Tyr
                        365   370   375
Lys Phe Ile Ser Lys Ala Lys Lys His Gly Ser Lys Cys Leu Val
                380   385   390
His Cys Lys Met Gly Val Ser Arg Ser Ala Ser Thr Val Ile Ala
                        395   400   405
Tyr Ala Met Lys Glu Tyr Gly Trp Asn Leu Asp Arg Ala Ser Ser
                410   415   420
Arg Ile Arg Leu Glu Lys Val Tyr Asn Leu Met Ser Phe Asp Val
                        425   430   435
Cys Thr Phe Pro Thr Ile Met Thr Ile Gln Ile Met Asn Ile Ser
                440   445   450
Ile Thr Pro Lys Lys Phe Pro Tyr Ala Ser Leu Ser Phe Ser Pro
                        455   460   465
Leu Ser Pro His Ser Ser His Ile His Arg Gln Ser Trp Ile Cys
                470   475   480
Phe Leu Leu Leu

<210> 10
<211> 1569
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1269536CB1

<400> 10
ccggaggaga ggcgaggagg aaacagtacc ggctgaaggc cggtcttgca ggacggggt 60
acggctggag gcggctttg gtggtgacg ctgtggggc gggggctggg gcttagagg 120
gtggtcttg gagatggag ggtcgaggct gctcgttgc ctggattctg gttccgccc 180
tggagcaagc cggggctgg tcggcagctg ggcggccatg gagtccacgc tgtgcgcggg 240
catcgtgata gggaggcgc tacagaacca gctagactgg ctggagaacg tgttgctctg 300
gatcaccttt ctgggtcatc ccaagatcct cttttgtttc tactacccg cttcctacta 360
cgcctcccgc cgtgtggca tcgcggttgct ctggatcagc ctcatcaacg agtggtcaa 420
cctcatcttc aagtggttc tttttggaga cagccctt tggtgggtcc atgagtctgg 480
ttactacagc caggctccag cccaggttca ccagttcccc tcttcttgtg agactggtcc 540

aggcagccct tctgggacact gcatgatcac aggagcagac ctctgggcca taatgacggc 600
cctgtcttct cagttggcca ctcctcccg cagcctctca gtaacgtca tgcctagcct 660
gccttattgc aactthcttt tgcccgttag nttgtgcaa actthatct tagcagactt 720
cctcaacag gtgttggttg gcctaataac tgtcggctgtc ctgcctgcn tcatgactcc 780
ntgacct atggaggggg agtaagtt ctatgggttg actgactgg cctcatgtt 840
aggcagaga ctcatctatt gacccctctt tacactgggc ctgtcctt cttggtcat 900
agcctagcc ttcaagtggt gtgacccc ctgagtgata cactggata gtcggccct 960
tgcctccgt agctttgact cagggtctcc attcgctg cctcatgcct tgaatttca 1020
ctngctatc tagttgctt cggcacagct gggaaatgg cagaagatac cctgcctct 1080
ctgggcat ggggctgatgg gccccttgga ctggctgggc cactccct agacgacgc 1140
tcctctacat tttctttt ctcaagtaca cctctgcc atgtctagtc ctggcctg 1200
tgccctggc agtccacatg ttcagtgcc aggaatcacc gccatcaa tctcctgac 1260
ttcttgtgt cctcccttc cttcatct aacaaagcaa acattctgtc aacacacac 1320
tccaggagg agtccaatc actccagc actaagtagg ccttttctc ctaaatttg 1380
cttctggacc acctggcctt agcccaaaa atgggcttt tctctccag ataagttggt 1440
cctcctctg cctttcctct caaggcccca aagagcaaag gcaacacacac gcccagcggg 1500
ttcttgcaac agtcttgagg gtatgcacgg cgtctccaat aaagccctto aatacttga 1560
aaaaaaaaa   1569

<210> 11
<211> 1158
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1849177CB1

<400> 11
tgccgtccct ctgcctgtgtc acacaatgg cccactgtgc cgcagcggac gtcaccaccc 60
gcctcgatga ggtgttctgg tttggagact tcaacttccg cctgagtggc gggggcacag 120
tcgtgacgc cctcctgtgc cagggcctgg tggtggactt gccgggcctg ctgcagcacg 180
accagctcat tcgggagatg cgggaaaggt ccattttcaa gggcttccag gagccggaca 240
tccactcct cccatcatac aagtttgaca tcgggaagga cacgtacgac aggacctcca 300
agcagagga ggcctcatac agggaccgcg tcttgtacag aagccgccac aagggcgaca 360
tctgtcctgt gagctactgt tcctgcccgg ggatcaagag gtccgaccac cggcctgtgt 420
atggcctctt ccgggtgaaa gtgaggccgg ggcgagacaa cattccgttg gcagctggca 480
aatttgatag agaactgtac ttactaggaa ttaaaagacg gattccgaag gagattccaga 540
gggagcaagc actacagagt cagaactca gaccatcg ctccgtttct tgaagtttgc 600
traacgaga ctcacagcty caagtgggg tgattgtatt gatcaaaac cactggaag 660
gacaaagagt ttgccgcctt tcggggatcc aagggactgt gggacacgty cctctgtgtc 720
acgtccag gaaggaagcc aacctgagc gagcctgtcc tctgtggcag gtccacacgg 780
tgtgggtggg caggggttgg accccgtcct ccatggcagg tccatacac atgggtggca 840
gggtttggac cgccaagta gacccacgga cccagcccac tctgggggc agacgtcaga 900
atccgttcct gagagctcc tgtgcctgg gggcagtcac agagccccc aaacccccg 960
tgcctgtgac cagcctctc ctccacacc gaagcaggcg tcatctgtg tcctcctgtc 1020
agccctcaaa acacacaca cccatgtca cagggtatc cagagcaagg ctggatctgg 1080
gctaggtgac cctgggggcc tcagacctct gagtggcccc cagagcccag ctgaggaagc 1140
cacacagcct ccaggacg   1158

<210> 12
<211> 111.
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3402521CB1

<400> 12
ctgttctgga tactctagta caaacaatca gtacagagag ccgtgctcag cggcctctca 60

```
actgggttgc tggacagacc cagccacggc ttacacacag caaacctggc cgattcaggc 480
ttcctggttg tcaagggtgg ttgaagtcgtg caccgatcag atgacacagca gcattattc 540
aacactccat tccagctctc aatcgctgcc cctcaagcgc aggaactgct cttgacgac 600
agtcacggatc ctgctgtatag cacgtctcct gatctcacagc tagcgagacat catcctcact 660
gcaacagatc gactctttga caacatcct gattatatga tcctccacatga cctaaaaacc 720
ttaaagaatt caaattatga cagtatacaa cacactccaa caagcacatcgc tgacccaagcc 780
catcagctcgc cctacgacct aaatcatatg tccctctttg cacacttttgg atgtgacaat 840
gcacttgacct tgacacgtcgg aaagccacat gacactacacg ctcttcttct cctactcggct 900
gactatacag actacttctaag gttcaagt ctgcctctcc tctccactcatc tccaaattcct 960
cctgcacact gtgctgatcc tcgtcgacga accacatcaca cttcccacatc atctccaatgc 1020
ccagtgatgc aagtcttcctc cttgcctctc tgacactagt tcacacatctt cttcacagacac 1080
actattacaa ttcactagtt catacctgcc                                   1116
```

```
<210> 13
<211> 1837
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1723447CB1

<400> 13
atgtccgccg gctgggttcg gcgccgcctc ctgcctgggg aggcgctccc cgcggccgtg 60
ccgcctgggc cgcatgccag cccggtgccc caccgacggc cccgcttcct tcgcgggcct 120
agctccagcc ccgggcgcgc cgaggcctcg cggccgccag actcccggcc cgtgcgcagc 180
cccgcacgag gacgcacgcc acctggaat gcaggctacg ccgagattat caatgcagag 240
aaatctgaat caatgagga tcaagccgc tgtgggaagg tgtgcatcg gagatgtgag 300
tttgtggttg aagaagagtg gctgacccctg tgcccagagg agtgcctgac aggccattac 360
tgggcagctgt ccgatggca cggcggtcct gcagcagcca tcctcggctgc caagacccct 420
cactcctgct tgcgccggca gctggaggcc gtgttgaag gcttgggtgc cactcagcca 480
cccatgcctc tcaatggccg cctgcatctgc cccagtgacc ctcagtttgt gggaggaaaag 540
ggcatcaggg cagaagactt ggtgatcggg gcattggaga gtgccttttca ggaatgtgat 600
gaggtgatcg ggcggagtcc gtgaggcctca ggcacagatgc gcggtctgacc agccctggtg 660
gctgtgtcc tgcagggaaa gctgtacatg gacaaatgtg gggataggag ggccatcttg 720
gttgcggatag atgacatacg gccactgagc ttcgagttca cccagagac tgacgcgcag 780
ggatctccgc agctggcctt tgtctatcct gagttccttgg ctggtgagtt cacccgactg 840
aagttcctc cgcgggatgac gggggatcac ttgggacaga aggttttgtt cagggatcac 900
cacatgagtg gctggagcta caaagtgtg gagaaatgg atctcaagta cccactgatc 960
catggcagg gtaggcaggc tggttacta ggaacactgg ctgtctcccg ggcctgggga 1020
gaccatcagg tcaagtcct cggacacaaac atcacagctca agccattctt gctctcctgtg 1080
caacagttga ctgtgctgaa tgtggaccag ctggagctac aggaggatga tgttggttgtc 1140
atgggaatcg atggacttcg gggatgtacg tccaacgagc aggtggcatg gctggtcggg 1200
agcttcctcc ctgggaacca agaggaaccc acacaggttc tcaaagccgg cccagatgct 1260
aatacacacg acacacaggaa aggaaagacag tctcacagag gaagggcagg tgtcctacga 1320
tgacgtctcc gtgttggtga ttcccttgca cagtcagggc caagagacga gtgaccactg 1380
aggattcaga cactgtatcc cagaactgct ctagtccgg ggtgtggbct gggcatcct 1440
ccagtgtcac caagagcaaa tcctcctgtc cctatccta gccacgccc agtgctctca 1500
ctatccacct caacacacat ccatctcaag aggaacatt atacccaggca gtcacagctg 1560
gaagtgtctg gagagcccag cccaccaggt cctgcctttt gcgtgataa cctcctctgg 1620
cagagtgact ttacaactta actaggaaac ccatgtgagg ctcctcagac aggatcttga 1680
anagcccaaa gtatcattct cagatggg caccaagtg aaggttatta gccaaagatg 1740
ccaggatggt tagctaccc atgtttagat ccaggtctcc aattcatggt tatcagggca 1800
tgtcttccca accacaaag tccacgcagg tggctga                          1837
```

```
<210> 14
<211> 1704
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1873677CB1

<400> 14
ttccgggtcg cgctaggccg gcgttgcgcc ggttgtgccg catctagaga gtgggggagc 60
cgccccgca cccacgccc ttcgcgctgc ctggtcgctg gtgaagccg cggcgcgcgc 120
```

```
rtctttaggg  cctgaagga  tttttcttcc  tggctggtt  ccttgcctt  cttcattttgc 180
tttattcgt  tggattttca  agaagcttty  aagctggata  ggagtctcca  gtcttcctct 240
tgatccatat  gtttcaggta  aaagaatgtc  acatgtcagc  atttgtacct  gaagtcagca 300
tgcaaagttc  agggtacatc  gatgaatgcc  aactttttgaa  ttccatgtt  gtattttgtg 360
accattctat  ctcggaacat  ccttaaaaga  cttcatgtat  cttactcagg  acacctgatc 420
ttttgagct  tcatacttca  catctagatg  tcacggtct  ttccatgtt  aacagttctg 480
acacatgtttt  attatatatg  cctttgggtgc  cgagtcagga  cagctacaag  aggagaaatg 540
atgaacacc  atagagctat  agaatcaaac  agccagactt  cccctctcaa  ttcagaggta 600
gtccagtatg  caaaagaagt  agtggatttc  agttccatt  atggaagtca  gaatactat 660
tcctatacta  tgtggaattt  ggctggtgta  caaaatgtat  tccaaagttc  tttgacttt 720
actccagaag  ctgtgtgttg  aacttcatgg  acatggtggg  attagtgtcc  cagtgctcc 780
ttgccattca  atgaggacgc  acctaattc  cagagcaagg  actatgtgga  acttactttt 840
gaacaacagg  tgtatcctac  agctgtacat  cttcagaaaa  cctatcatcc  cgtagcagtc 900
attagaattc  tccttgtc  tgcaaatcc  tattccaa  atccacagc  tcaagtaaga 960
tcggaatt  tctggttaaga  gagacctacg  aaggtgaaty  cttcccaagc  tcgccagtt 1020
aaacttgta  ttaagcagat  aaatttcccc  acaaaatcta  tacgactgga  agtaaatagt 1080
tctcttctgg  aatattacac  tgaattacat  gcagttgtgc  tacatggtgt  gaaggacaaa 1140
ccagtgcttt  ctctcaagac  ttcacttatt  gacatgaatg  acatatagaaga  ccatgccat 1200
ccagaaaagg  atggttgtga  aatggacagt  cttaacaaaa  agtttagcag  cgtgtgtcctc 1260
gggaaggg  caaataatgg  gtatttttgat  aaaatacctt  atgagtaag  ccaaaaatat 1320
ttaggagcag  tattggatat  aacactatag  aattttaatc  tttagatatt  aatgtttact 1380
gagcatcggt  aattgagcat  tggtaagtta  tatataattc  gatgtgtttgt  atattttgat 1440
tttggaaaaa  ctagacaaaa  tcatgtttta  acaattttca  ctgctttgt  tcttggtttgt 1590
catt                                                                1594
```

```
<210> 15
<211> 1407
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6205333CB1

<400> 15
```

```
ggctacggaa  gcggtgagac  tgtctctcgg  ctgcagccct  ggtgtgaccc  ggcctgttcg 60
cgtagagaty  ggcaggtctg  gatggagtgg  ggtgcggtga  gctgagctga  ccctgcttcg 120
caacggggac  tgcagtgacc  ccggcttgcc  ggcagggcgg  gtaaacaggtt  gagccagggt 180
ggggctgctg  agggggcggg  aggattctc  tgaagacccg  gcaacaggcta 240
ttcctttctg  cgacgacgcc  attgctatgg  aaacaaaagc  gttaggcag  cgggggattga 360
ggctgcggga  tcatgacggg  tctctctcc  gaagaaccct  gcctaaggct  tcccaagcg 360
gctacttct  gagcgaaccc  gccaccggc  ctgaaggaga  gagttttca  tgacaccagc 420
ctagcagtaa  gacgcagcct  ccgtgcttcg  ctgactgctg  accactgacc  cacgggcttg 480
atgacagcac  cctcgtgtgc  cttccccagtt  cagttccggc  agccctcagt  cagggccctc 540
tggcagataa  ccaaaagcct  gtatatcagc  aatgtgtgtcg  ccgcaacaa  caagctcatg 600
ctgtctagca  aacaagatca  catggtcatc  aatgtgctag  tgtaggtagt  gaacacttcg 660
tatgaggata  tcgagtgct  gtgttgact  ccctaactg  acgtctctgt 720
gacttctttg  acccttattg  tgacccatatc  cacagcgtgg  agatgaagca  ggggcgtact 780
ttgctgcctct  gtgctgctgg  tgtgagcggc  tcagctgccc  tgtgcctgg  ctacctcatg 840
cagtacaac  caattccct  tctgcacgc  cacacgtcaa  caacatcatc  ccgtcccatc 900
atccacttta  atacgccct  ttggaacac  ttatccaacc  atcacttca  actcctggc 960
cagaaactgc  tgtacatgt  cacttcccac  ctgggacgca  tttctacacc  ctatgccag 1020
caagtccct  ttatatatc  cttgtgccc  atccacaag  ctcacacatc  gctgcagat 1080
ctacacttc  catactcat  tcaataac  atcaagctc  cacactcta  catttctgt 1140
tacacaaaa  aacaacatat  acgtttata  ctacctgtc  tcctcccctc  cacttttcct 1200
cttctaact  catttttct  tagtctaac  ttcctgtcc  atgtgcatc  acacttttcc 1260
gtcatgtag  tgtacctggc  cagggacta  gttgtggag  atgtgtccc  gaagaaagat 1320
ggcagcctca  ttgcacatgg  cagcgcttgaa  atcgtgcact  actccaaaa  acaagattgc 1380
ctaggaatga  tctcctacaa  ttcacc                                      1407
```

```
<210> 16
<211> 541
<212> DNA
```

<223> Incyte ID No: 7472032CB1

<400> 16
atgcaaaagc ggaaacttt gataaaacta tcagatctct tgagacttat tcgctactac 60
gagaacagta tggggggaaac tgcccccatg atccaaatta tctcgcacg ggtccctctc 120
cacactttg tgaattatgg gagtacaatt caagatgaga tttggataag agtatgtaaa 180
ggaatttatg acactactct tctggagaaa gaaggtaccc atggtctttt 240
gctgatgtg caacaccat caacacgatt gatgactggt taagttttgt caaaattaar 300
tttataaag aatttatg tgactgct gttcattgca ttgcaggctt tgggagagt 360
caagtacttg ttgcctag attaattgaa gggaaaatga aagatgatga ttgagtagaa 420
tttgcaaga aaaagcagta tggagcttt aacagcaagt taacttttgt attggaaga 480
gtatcatcct taaagatgtg gntgcattt aaagaccca gtggtcatag ataaaatgt 540
tcagttaa   549

<210> 17
<211> 1537
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2522707CB1

<400> 17
aggagctagg gttgtcctgc caggtagaa tggaaccgt aggaggggta cttaaacgga c 60
cggcctacca ggcctgtggc cgtcgcggg aagagcact cagatctcag gatgatgggg 111
cacagcttgg gtttgcatg cagcacatcg tggtgtgcc ccaagtacty gttcggaggg 171
gcttcttgg aaggggaccctc ttcatgacca ggactctctg cagcccaggc ccaagccagc 241
tcggagagaa aagacctgag gagttgtgcc tggggctgca ccacggctc cccagcacgg 300
gaagagccct gggggacagc attcagcaac gagcgacctc cacagccaag acttggttgg 360
acagatatga agagttgtt ggactcaacg aggttcgaga gtccccagcga aaggtgacag 420
aggctgagaa agttgttcatg gttggctcgag ggcctgtccg agaggctcgg gaggacttgg 480
aagttcatcca ggccaagctg aaggaggtga ggggacgctt ggaccgtgtc tccagggagg 540
acagtcagta cttggaacty gctactctcg agacacggat gctgcaggag gagaagaggc 600
tttgcacacc ctatctcgt gcagaaagact ctgagcgaga gaagttcctcc ctcttctctg 660
aagctgtgcg ggaaagtcat gagaaggagc gcacaaggge tgagaggacc aagaactggt 720
cactcattgg ctcagtcctg ggggccctga ttggtgtggc tggctccacc tatgtgaacc 780
gtgtgcgact acaggtcgtg aaggcttta tcctgcaggc gcagaaggg cctgtgagtc 840
tcaagaggc cattcgagaa cagggcgtcta gctactccg ccagcaggac gacctccaca 900
attttatgt ggacttgagg tgcctggtac atgtcgctgg gccagggcag gacctggt 960
gcagggcagg tagtccccg accagagaca gagatgtaga tgtcctttca gctgcccttga 1020
gagacagct tagtcattcc aggcaagtcc attcatgctct agaaggctta cgagagcagc 1080
ttgatggcc agaaaagact tgtagccaaa tggctgggt ggttcaagctt gtaaagtctg 1140
cagcacacc aggcctggtg gaaccagcag aggggctat gcccagcttc ttgctggagc 1200
aagttagcat catctgggca ctgtcagaca cggagcagag actagaagcc caagtcaaca 1260
gcaaaccat ctatgacac ctggtcacct gtgacactt tgtggccaca cgctgtgtgc 1320
tctaatatt attcaaagct agctaaatccc tggcctctcc tccagggggt ctgaggcaat 1380
agctgtgaat gtggatttaa gtagagaatc gtagcaatga agcgagcctt tgggggcatg 1440
tacaacctca atctgaagga gcagtatctg tgtgggtcac cagcaggcat gctccgcctt 1500
gtagacaagg ttcatttaca ttaattatca aaactttgtg ctaatgtcca attaaaatat 1560
catgagtttt attatttaa aaaaaaa   1587

<210> 18
<211> 1771
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 637331CB1

<400> 18
gtggaagacc cagcagcgcc tagatggctt tggtcacggt cagggtca cctaccccga 60
gtaccacctc cagccctcgc ggctggagg cagagagtgg ggaggaagaa tgcgggtgac 120
ggtagggag catcagcgag agctttctaa ctgtcaaagg tgctgcctt tttctaccac 180
ggggaaatgg ctcatccaca ccaagaatca gccacagacg gaacaagcat gcagggatc 240
tccaacagca tctccaagca atgttcattt tactccgccc agaagacaac atcaggctgg 300

```
ttgtaagatt ggaaagtatt tataagtatt aaatangtta tatgstattg gtttttatta  360
atggtacata ttaaatttaa gaatgtattg ttttaggaat ggattttttt tttaatgata  420
gtaaaattg taaaatggtt ttagttttgt ttttttggag tgaaaagtta attatttgg  480
atggtgatgg tgggttttagt gtatggacgg ataaacagagt ttatatattc aaacttgtat  540
ttgtgtaggc aatgtgggtct gcaatacaga gcttacacaa ggcttgtgaa gtcgccagag  600
tgcataacta ctaccatgc atgctatttc tcaottttttt gagttattat gagagccata  660
tcaactcaga tcaatcctca gtcaatgaat ttaatttaat gtaagatgta cagtttttttt  720
tgggtgattc tttggttttt ttttatttgata tatttattta atgttgaatta aatataattt  780
taattaaaac aaaattaatt gatattcatga ttttaggagta tttggagaat attttattttt  840
aatagataag aatagagttg gaaatgtaaa ttgtgtgtttt cttgtggggat tttaaggaat  900
ttatagataa ttgaaatgata gtgatctttg gttaaattgtt ttgttttttt cagatatttg  960
atcatgtgtt cctggggtca gaatggaatg tttccaaott agaggactta caggaacgag 1020
gtggtacggta tatcttgaat gttaactgag atatagataa cttttttccca gtggttttttg 1080
aatatcataa cattgggtta tatgatgaag atgcaatttta ttttcctggcg tactggaatg 1140
atatttataa attatttttt gaagtaaaga aatatggata tatatgcttt gtgtattgta 1200
aaatgtttgt gagttggttca gtttttaacg ttattgccta tgtaatgaag gaatatggct 1260
gggaatttgga cggacctcg atccgaattc ggctcgagtaa agtgtacaat ttgatgagtt 1320
ttgactgtg cacctttcca accatcatga caatacaaat aatgaacata tcaatcaccc 1380
caaaaaagtt tcctttatcc tcttttgtcat tcagccccct ttcccacat tctttccaca 1440
tccacaggta atcatggatc tgttttctat tgtttattaat tagtttgcat ttcctacaat 1500
tttatacaca ttgaataaaa ctataggtac tcttgtttag ctcttttcat tcagccaat 1560
aagtttgaga ttcattcatg ttgttatatg cgtctatgtt cattccttt tattgctaag 1620
tgttattttca ttatatggac ataccagaat ttgtttatcc attcatgttt tgatggaaat 1680
ttgtattttc agtttttgc tattacaaat aaaactgatg tcaacatttg tttacataca 1740
atatgaaaga aaagaaagaa cgaaaaaaaa g                                  1771
```